Subject: Re: Can DW handle big data?
Posted by thomas on Thu, 24 Oct 2019 18:59:06 GMT
View Forum Message <> Reply to Message

Dear DaRong,

3 million rows is already a lot. I recommend for very large files to use DataWarrior on Linux, because you can easily increase the memory maximum that there is at least no memory problem. DataWarrior uses multithreading for most functions, which benefit from it. However, reading a file is a serial process and cannot easily be parallelized. Possibly I could gain some performance, when distributing the data analysis after file loading on multiple cores. I will put it on the agenda, but not before the next release, which I anticipate before the end of the year.

Thanks and best wishes,

Thomas